

OUTLIERS IN THE SPECTRUM OF IID MATRICES WITH BOUNDED RANK PERTURBATIONS

TERENCE TAO

ABSTRACT. It is known that if one perturbs a large iid random matrix by a bounded rank error, then the majority of the eigenvalues will remain distributed according to the circular law. However, the bounded rank perturbation may also create one or more outlier eigenvalues. We show that if the perturbation is small, then the outlier eigenvalues are created next to the outlier eigenvalues of the bounded rank perturbation; but if the perturbation is large, then many more outliers can be created, and their law is governed by the zeroes of a random Laurent series with Gaussian coefficients. On the other hand, these outliers may be eliminated by enforcing a row sum condition on the final matrix.

1. INTRODUCTION

This paper is concerned with the study of outliers of the *circular law* for iid random matrices and its variants. To recall this law, we make some definitions:

Definition 1.1 (iid random matrix). An *iid random matrix* is an $n \times n$ random matrix $X_n = (x_{ij})_{1 \leq i, j \leq n}$ (or more precisely, a nested sequence X_1, X_2, \dots of such matrices) whose entries x_{ij} for $i, j \geq 1$ are independent identically distributed complex entries, which we normalise to have mean zero and variance one. We say that such a matrix has *atom distribution* x if all the x_{ij} have distribution x , thus $\mathbf{E}x = 0$ and $\mathbf{E}|x|^2 = 1$.

Definition 1.2 (ESD). Given an $n \times n$ complex matrix A_n (not necessarily Hermitian or normal), we define the *empirical spectral distribution* μ_{A_n} of A_n to be the probability measure

$$\mu_{A_n} := \frac{1}{n} \sum_{j=1}^n \delta_{\lambda_j}$$

where $\lambda_j = \lambda_j(A_n)$ for $j = 1, \dots, n$ are the eigenvalues of A_n (counting multiplicity, and ordered arbitrarily).

If A_n is a random $n \times n$ complex matrix (so that μ_{A_n} is also random), we say that μ_{A_n} converges *in probability* (resp. *almost surely*) to another (Borel) probability measure μ on the complex plane \mathbb{C} if for every smooth, compactly supported function $F : \mathbb{C} \rightarrow \mathbb{C}$, $\int_{\mathbb{C}} F d\mu_{A_n}$ converges in probability (resp. almost surely) to $\int_{\mathbb{C}} F d\mu$.

The following theorem is the culmination of the work of many authors [23], [33], [24], [4], [6], [26], [25], [34], [42], [27], [44]:

Theorem 1.3 (Circular law for iid matrices). *Let X_n be an iid random matrix. Then $\mu_{\frac{1}{\sqrt{n}}X_n}$ converges almost surely (and hence also in probability) to the circular measure μ_c , where $d\mu_c := \frac{1}{\pi} 1_{|z| \leq 1} dz$.*

The result as stated is [44, Theorem 1.15], but this result is based on a large number of partial results (in which more hypotheses are placed on the atom distribution x) which are proven in the previously cited papers.

T. Tao is supported by a grant from the MacArthur Foundation, by NSF grant DMS-0649473, and by the NSF Waterman award.

The circular law implies in particular that the spectral radius

$$\rho\left(\frac{1}{\sqrt{n}}X_n\right) = \lim_{k \rightarrow \infty} \left\| \left(\frac{1}{\sqrt{n}}X_n\right)^k \right\|_{op}^{1/k} = \sup_{1 \leq j \leq n} \left| \frac{1}{\sqrt{n}}\lambda_j(X_n) \right|$$

is at least $1 - o(1)$ almost surely, where $o(1)$ goes to zero as $n \rightarrow \infty$. When the atom distribution x has finite fourth moment, we in fact have an asymptotic for the spectral radius:

Theorem 1.4 (No outliers for iid matrices). *Let X_n be an iid random matrix whose atom distribution x has finite fourth moment: $\mathbf{E}|x|^4 < \infty$. Then $\rho(\frac{1}{\sqrt{n}}X_n)$ converges to 1 almost surely (and hence also in probability) as $n \rightarrow \infty$. In fact, for any finite $m \geq 1$, $\|(\frac{1}{\sqrt{n}}X_n)^m\|_{op}$ converges to $m + 1$ almost surely (and hence also in probability) as $n \rightarrow \infty$.*

Furthermore, if all moments of x are finite, one has the tail bound

$$\left\| \left(\frac{1}{\sqrt{n}}X_n\right)^m \right\|_{op} \leq m + 1 + \varepsilon$$

with overwhelming probability¹ for each $\varepsilon > 0$.

Proof. This follows from what is by now a routine application of the truncation method and the moment method; see [10], [22], or [9, Theorem 5.17]. We remark that the tail bound can also be deduced from the main result using the Talagrand concentration inequality (after first truncating to the case when x is bounded); see [28], [1], [32]. The precise expression $m + 1$ is not important for our arguments here; any quantity that was subexponential in m would have sufficed. \square

Informally, Theorem 1.4 asserts that when the fourth moment is finite, there are no significant *outliers* to the circular law: with probability² $1 - o(1)$, all of the eigenvalues of the matrix $\frac{1}{\sqrt{n}}X_n$ lie within $o(1)$ of the support $\{z \in \mathbb{C} : |z| \leq 1\}$ of the circular law. The fourth moment condition here is necessary for the second conclusion of Theorem 1.4 (see [9]), and it is very likely that it is also necessary for the first conclusion.

Now we consider the circular law and its outliers for random matrices formed as a low-rank perturbation of an iid random matrix. The circular law is stable under such perturbations:

Theorem 1.5 (Circular law for low rank perturbations of iid matrices). [44, Corollary 1.17] *Let X_n be an iid random matrix, and for each n , let C_n be a deterministic matrix with rank $o(n)$ obeying the Frobenius norm bound*

$$\|C_n\|_F := (\text{trace } C_n C_n^*)^{1/2} = O(n^{1/2}).$$

Then $\mu_{\frac{1}{\sqrt{n}}X_n + C_n}$ converges both in probability and in the almost sure sense to the circular measure μ_c .

Remark 1.6. Thanks to a recent result of Bordenave [16], the $O(n^{1/2})$ bound here can be relaxed to $O(n^{O(1)})$.

However, the low rank perturbation C_n can now create outliers. Our first main result is to describe these outliers in the case when C_n has bounded rank and bounded operator norm, and x has finite fourth moment. In this case, it turns out that the outliers of $\frac{1}{\sqrt{n}}X_n + C_n$ are close to those of C_n . More precisely, we have

Theorem 1.7 (Outliers for small low rank perturbations of iid matrices). *Let X_n be an iid random matrix whose atom distribution has finite fourth moment, and for each n , let C_n be a deterministic matrix with rank $O(1)$ and operator norm $O(1)$. Let $\varepsilon > 0$, and suppose that for all sufficiently large n , there are no eigenvalues of C_n in the band $\{z \in \mathbb{C} : 1 + \varepsilon < |z| < 1 + 3\varepsilon\}$, and there are j eigenvalues $\lambda_1(C_n), \dots, \lambda_j(C_n)$ for some $j = O(1)$ in the region $\{z \in \mathbb{C} : |z| \geq 1 + 3\varepsilon\}$. Then, almost surely, for sufficiently large n , there are*

¹We say that an event E_n depending on n occurs with *overwhelming probability* if for every $A > 0$ there exists $C > 0$ such that $\mathbf{P}(E_n) \geq 1 - Cn^{-A}$ for all n .

²The asymptotic notation $O(), o()$ that we use here will be defined in Section 1.3.

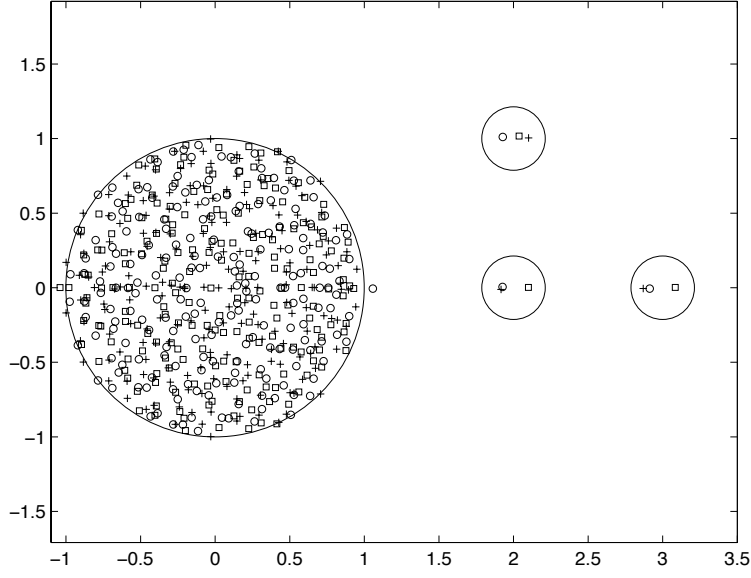


FIGURE 1. This figure shows the eigenvalues of three 200 by 200 iid random matrices with atom distribution x defined by $\mathbf{P}(x = 1) = \mathbf{P}(x = -1) = 1/2$, each of which was perturbed by adding the matrix $\text{diag}(2 + i, 3, 2, 0, 0, \dots, 0)$. The small circles are centered at $2 + i$, 2, and 3, respectively, and each have radius $n^{-1/4}$ where $n = 200$. (Figure by Phillip Wood.)

precisely j eigenvalues $\lambda_1(\frac{1}{\sqrt{n}}X_n + C_n), \dots, \lambda_j(\frac{1}{\sqrt{n}}X_n + C_n)$ of $\frac{1}{\sqrt{n}}X_n + C_n$ in the region $\{z \in \mathbb{C} : |z| \geq 1 + 2\varepsilon\}$, and after labeling these eigenvalues properly, $\lambda_i(\frac{1}{\sqrt{n}}X_n + C_n) = \lambda_i(C_n) + o(1)$ as $n \rightarrow \infty$ for each $1 \leq i \leq j$.

Thus, for instance, if one perturbs $\frac{1}{\sqrt{n}}X_n$ by a bounded rank, bounded operator norm matrix C_n whose eigenvalues all lie inside the unit disk $\mathbf{D} := \{z \in \mathbb{C} : |z| \leq 1\}$ (e.g. C_n could be a nilpotent matrix), then no outliers are created; but once C_n has eigenvalues leaving the unit disk, the perturbed matrix $\frac{1}{\sqrt{n}}X_n + C_n$ will also have outliers in asymptotically the same location. Theorem 1.7 is illustrated in Figures 1, 2.

Remark 1.8. An analogous result for Wigner matrices instead of iid matrices has recently been established in [18], [37], with more precise control (in particular, a central limit theorem) on the distribution of the outlier eigenvalues; the methods used are somewhat different, but the techniques developed here can be adapted to the Wigner case (Alexander Soshnikov, private communication). See also [21], [13] for further results in the Wigner case, whose methods are close to those used here, [35] for a treatment of the GUE case, [11] for a treatment of the LUE case, and [5], [12], [7] for a treatment of the covariance matrix case. Interestingly, in the Wigner case the outlier eigenvalues $\lambda_i(\frac{1}{\sqrt{n}}X_n + C_n)$ of the perturbed matrix are not close to the outlier eigenvalues $\lambda_i(C_n)$ of the original matrix, but rather to the shifted eigenvalues $\lambda_i(C_n) + \frac{\sigma^2}{\lambda_i(C_n)}$, where σ^2 is the variance of the entries of the Wigner matrix σ . This is ultimately because the powers $(\frac{1}{\sqrt{n}}X_n)^m$ have a significant presence on the diagonal in the Wigner case, in contrast with the iid case where all entries are small. Alternatively: the Wigner semicircular law has nonzero moments, while all nontrivial (pure) moments of the circular law vanish.

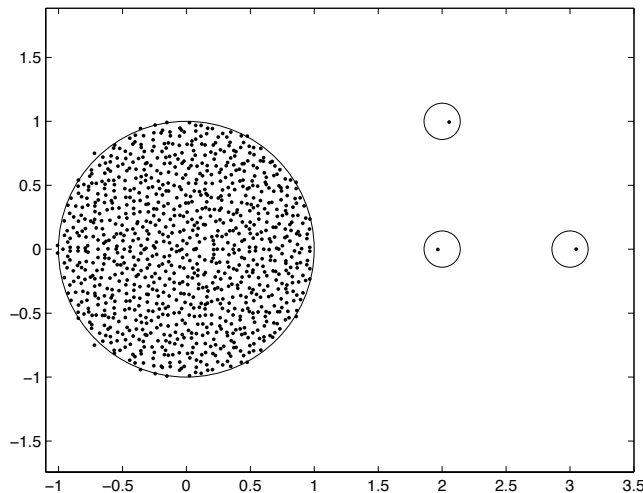


FIGURE 2. This figure shows the eigenvalues of a single 1000 by 1000 iid random matrix with atom distribution x defined by $\mathbf{P}(x = 1) = \mathbf{P}(x = -1) = 1/2$ which was perturbed by adding the matrix $\text{diag}(2 + i, 3, 2, 0, 0, \dots, 0)$. The small circles are centered at $2 + i$, 2 , and 3 , respectively, and each have radius $n^{-1/4}$ where $n = 1000$. (Figure by Phillip Wood.)

Theorem 1.7 is proven in Section 2. The main tools are asymptotics of Stieltjes transforms outside of the unit disk \mathbf{D} , combined with the fundamental matrix identity³

$$(1) \quad \det(1 + AB) = \det(1 + BA)$$

valid for arbitrary $n \times k$ matrices A and $k \times n$ matrices B . Note that the left-hand side is an $n \times n$ determinant, while the right-hand side is a $k \times k$ determinant. For low rank perturbations, we will be able to apply (1) with k bounded and n going to infinity, allowing one to transform an unbounded-dimensional problem into a finite-dimensional one.

Theorem 1.7 only deals with perturbations that are relatively small, having an operator norm of $O(1)$. It is also of interest to consider larger perturbations, such as those caused by adjusting the mean of each coefficient of X_n by $O(1)$. Here, the situation is more complicated, and we will consider only a few model perturbations, rather than attempt to obtain the most general result.

We first consider the case of iid matrices with non-zero mean, which we write as $\frac{1}{\sqrt{n}}X_n + \mu\sqrt{n}\phi_n\phi_n^*$, where μ is a fixed complex number (independent of n) and ϕ_n is the unit column vector $\phi_n := \frac{1}{\sqrt{n}}(1, \dots, 1)^*$; this corresponds to shifting the atom distribution x by μ (so that it has mean μ rather than mean zero). This is a rank one perturbation of $\frac{1}{\sqrt{n}}X_n$. The circular law still holds for this ensemble, thanks to Theorem 1.5 (or the earlier result of Chäfai [19]). However, in view of Theorem 1.7, we expect a single large outlier near $\mu\sqrt{n}$. This is indeed the case:

Theorem 1.9 (Outlier for iid matrices with nonzero mean). *Let X_n be an iid random matrix whose atom distribution has finite fourth moment, and let $\mu \in \mathbb{C}$ be a non-zero quantity independent of n . Then almost surely, for sufficiently large n , all the eigenvalues of $\frac{1}{\sqrt{n}}X_n + \mu\sqrt{n}\phi_n\phi_n^*$ lie in the disk $\{z \in \mathbb{C} : |z| \leq 1 + o(1)\}$, with a single exception taking the value $\mu\sqrt{n} + o(1)$.*

³We thank Percy Deift for emphasising the importance of this identity in random matrix theory.

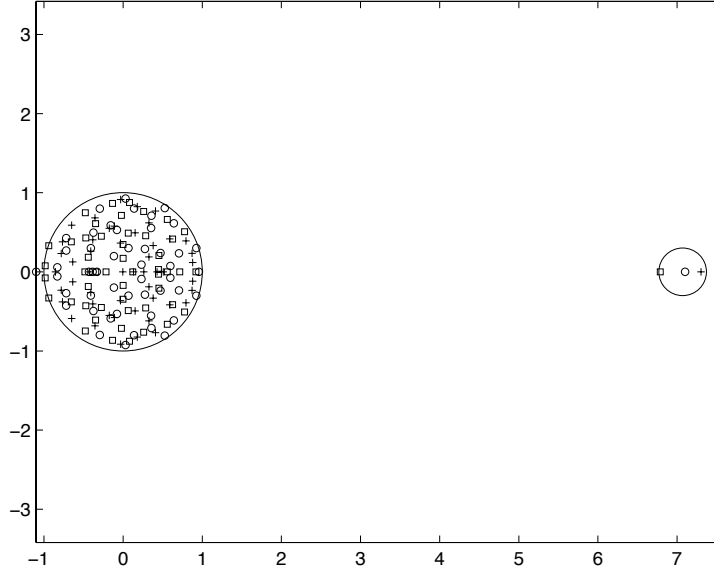


FIGURE 3. This figure shows the eigenvalues of three 50 by 50 iid random matrices with atom distribution x defined by $\mathbf{P}(x = 1) = \mathbf{P}(x = -1) = 1/2$, each of which was perturbed by adding the matrix $\mu\sqrt{n}\phi_n\phi_n^*$ where $\mu = 1$. The small circle is centered at $\sqrt{50}$ and has radius $n^{-1/4}$ where $n = 50$. (Figure by Phillip Wood.)

We prove this result in Section 3. One can obtain more precise information on the distribution of this exceptional eigenvalue, particularly if one assumes more moment hypotheses on the atom distribution x ; see [41]. The existence of this exceptional eigenvalue was already noted back in [2]. Theorem 1.9 is illustrated in Figure 3. Figure 4 corresponds to the case of a smaller value of μ , and falls instead under the regime covered by Theorem 1.7.

Next, we consider a model that was introduced in [39], in the context of neural networks. In our notation, this model takes the form

$$(2) \quad A_n := \frac{1}{\sqrt{n}}X_n + \mu B_n$$

where $\mu > 0$ is a fixed parameter, and B_n is a random matrix (independent of X_n) such that the columns of B_n are iid, with each column equal to $\sqrt{\frac{1-p}{p}}\phi_n$ with probability p and $-\sqrt{\frac{p}{1-p}}\phi_n$ with probability $1-p$, for some fixed $0 < p < 1$. (In the notation of [39], the excitatory mean μ_E is $\mu\sqrt{(1-p)/p}$ and the inhibitory mean μ_I is $-\mu\sqrt{p/(1-p)}$, and the excitatory and inhibitory variances are assumed to be equal.) Note that one can write

$$(3) \quad A_n = \frac{1}{\sqrt{n}}X_n + \mu\sqrt{n}\phi_n\psi_n^*$$

where ψ_n is a random vector whose entries are iid and equal $\frac{1}{\sqrt{n}}\sqrt{\frac{1-p}{p}}$ with probability p and $-\frac{1}{\sqrt{n}}\sqrt{\frac{p}{1-p}}$ with probability $1-p$; with this normalisation, ψ_n has mean zero and unit variance.

Again, by Theorem 1.5, the ESD of A_n is governed by the circular distribution μ_c in the limit $n \rightarrow \infty$; however, as observed numerically in [39], a small number of outliers also appear for A_n .

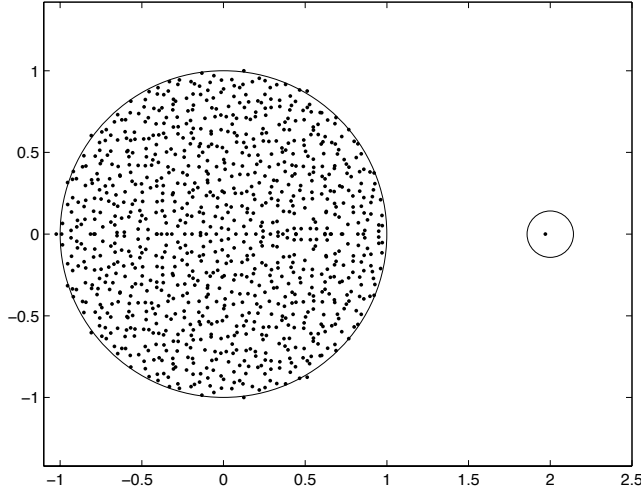


FIGURE 4. This figure shows the eigenvalues of a single 1000 by 1000 iid random matrix with atom distribution x defined by $\mathbf{P}(x = 1) = \mathbf{P}(x = -1) = 1/2$ which was perturbed by adding the matrix $\mu\sqrt{n}\phi_n\phi_n^*$ where $\mu = 2/\sqrt{1000}$. The small circle is centered at 2 and has radius $n^{-1/4}$ where $n = 1000$. (Figure by Phillip Wood.)

It is possible to explain the outliers by the arguments of this paper. However, in contrast to the situations in Theorem 1.7 or Theorem 1.9, in which the outliers essentially have a deterministic location up to $o(1)$ errors, for the model (2), the outliers retain significant randomness at macroscopic scales, and need to be modeled by a point process rather than by a deterministic law. To describe this point process, we introduce the k -point correlation functions $\rho_{A_n}^{(k)} : \mathbb{C}^k \rightarrow \mathbb{R}^+$ for the ESD of A_n for $1 \leq k \leq n$, defined as the unique symmetric function such that

$$\int_{\mathbb{C}^k} F(z_1, \dots, z_k) \rho_{A_n}^{(k)}(z_1, \dots, z_k) d^2 z_1 \dots d^2 z_k = \frac{1}{k!} \mathbf{E} \sum_{i_1, \dots, i_k \in \{1, \dots, n\}, \text{ distinct}} F(\lambda_{i_1}(A_n), \dots, \lambda_{i_k}(A_n))$$

for all continuous, compactly supported test functions $F : \mathbb{C}^k \rightarrow \mathbb{C}$; note that the right-hand side is not dependent on how one orders the n eigenvalues of A_n . Here, $d^2 z$ denotes two-dimensional Lebesgue measure on \mathbb{C} . If the atom distribution of A_n is discrete, then $\rho_{A_n}^{(k)}$ needs to be interpreted as a distribution or measure rather than as a function, but this technicality will not concern us here.

It turns out that the correlation functions $\rho_{A_n}^{(k)}$ have a limiting law $\rho_\infty^{(k)}$ outside of the unit disk \mathbf{D} (inside the disk, one expects these functions to go to infinity, thanks to the circular law and the choice of normalisation). We do not have a completely explicit formula for this limit, but can describe it instead as the zeroes of a random Laurent series. More precisely, consider the random Laurent series

$$g(z) := 1 - \mu \sum_{j=1}^{\infty} \frac{g_j}{z^j}$$

where g_1, g_2, \dots are iid copies of the real Gaussian distribution $N(0, 1)_{\mathbb{R}}$. From the Borel-Cantelli lemma we see that this Laurent series is almost surely convergent in the complement $\mathbb{C} \setminus \mathbf{D}$ of the closed unit disk \mathbf{D} , and almost surely has a finite number of zeroes in the region $\{z \in \mathbb{C} : |z| > 1 + \varepsilon\}$ for any fixed $\varepsilon > 0$. We then define the limiting correlation function $\rho_\infty^{(k)} : (\mathbb{C} \setminus \mathbf{D})^k \rightarrow \mathbb{R}^+$ outside of this disk as the unique

symmetric function (or more precisely, distribution) such that

$$(4) \quad \int_{\mathbb{C}^k} F(z_1, \dots, z_k) \rho_{\infty}^{(k)}(z_1, \dots, z_k) d^2 z_1 \dots d^2 z_k = \frac{1}{k!} \mathbf{E} \sum_{w_1, \dots, w_k \in \Lambda_g, \text{ distinct}} F(w_1, \dots, w_k)$$

for all continuous, compactly supported test functions $F : (\mathbb{C} \setminus \mathbf{D})^k \rightarrow \mathbb{C}$, where Λ_g are the zeroes of g (counting multiplicity); a little more explicitly, $\rho_{\infty}^{(k)}(z_1, \dots, z_k)$ can be defined for distinct $z_1, \dots, z_k \in \mathbb{C} \setminus \mathbf{D}$ as

$$\rho_{\infty}^{(k)}(z_1, \dots, z_k) = \lim_{\varepsilon \rightarrow 0} \frac{\rho_{\infty}^{(k), \varepsilon}(z_1, \dots, z_k)}{(\pi \varepsilon^2)^k}$$

where $\rho_{\infty}^{(k), \varepsilon}(z_1, \dots, z_k)$ is the probability of the event that there is a zero of g within ε of z_j for each $j = 1, \dots, k$.

Remark 1.10. If the g_1, g_2, \dots were *complex* Gaussian $N(0, 1)_{\mathbb{C}}$ instead of real, we normalised $\mu = 1$, and we replaced the constant coefficient 1 by another complex gaussian g_0 , then $g(z)$ would be a Gaussian power series (GPS) in the variable $1/z$. Gaussian power series have been intensively studied (see the recent text [29] and the references therein). In that case, the k -point correlation functions are given by the determinantal formula

$$\rho_{GPS}^{(k)}(z_1, \dots, z_k) = \frac{1}{\pi^k} \det\left(\frac{1}{1 - z_i \bar{z}_j}\right)_{1 \leq i, j \leq k};$$

see [36]. One may then hope that a somewhat analogous formula might be obtained for the random Laurent series considered here, possibly using the explicit formulae for the correlation functions of zeroes of real random polynomials from [38] as a starting point. We will not pursue this matter.

We can now state our main theorem regarding this model, which we prove in Section 5:

Theorem 1.11 (Limiting law). *Let X_n be an iid random matrix whose atom distribution x is real-valued and which is either gaussian (i.e. $x \equiv N(0, 1)_{\mathbb{R}}$) or bounded, and let $0 < p < 1$ and $\mu > 0$ be fixed. Let A_n , $\rho_{A_n}^{(k)}$, and $\rho_{\infty}^{(k)}$ be defined as above.*

- (i) *(Crude upper bound) For any $\varepsilon > 0$, let N_{ε} denote the number of eigenvalues of A_n in the region $\{z \in \Omega : |z| \geq 1 + \varepsilon\}$ (counting multiplicity). Then $\sup_n \mathbf{E} N_{\varepsilon}^m < \infty$ for all $\varepsilon > 0$ and $m \geq 1$.*
- (ii) *(Limiting law) $\rho_{A_n}^{(k)}$ converges in the vague topology to $\rho_{\infty}^{(k)}$ on $(\mathbb{C} \setminus \mathbf{D})^k$. In other words, one has*

$$\int_{\mathbb{C}^k} F(z_1, \dots, z_k) \rho_{A_n}^{(k)}(z_1, \dots, z_k) d^2 z_1 \dots d^2 z_k \rightarrow \int_{\mathbb{C}^k} F(z_1, \dots, z_k) \rho_{\infty}^{(k)}(z_1, \dots, z_k) d^2 z_1 \dots d^2 z_k$$

whenever $F : (\mathbb{C} \setminus \mathbf{D})^k \rightarrow \mathbb{C}$ is continuous and compactly supported (in particular, it is supported in the region $\{(z_1, \dots, z_k) \in \mathbb{C}^k : |z_1|, \dots, |z_k| \geq 1 + \varepsilon\}$ for some $\varepsilon > 0$). In particular, the limiting distribution is universal with respect to the distribution x .

Remark 1.12. The requirement that all the coefficients of X_n are real is a natural one from the neural net application[39]. However, in view of the better developed theory for complex Gaussian power series[29], it may in fact be more natural from a theoretical perspective to consider the case when the X_n are complex valued, e.g. if the atom distribution is complex Gaussian. In that case, there is a similar result to Theorem 1.11 but with the coefficients g_j of the random Laurent series $g(z)$ given by complex Gaussians rather than real Gaussians; we omit the details. The requirement that x be either Gaussian or bounded is a technical one, so that one may apply concentration inequalities; it may certainly be relaxed substantially.

Theorem 1.11 is illustrated in Figure 5.

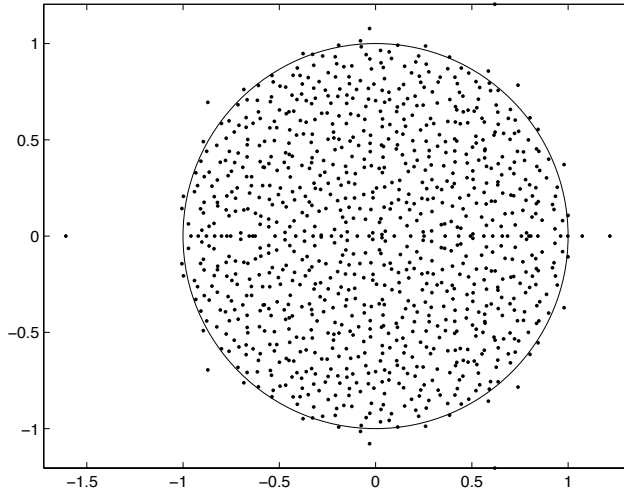


FIGURE 5. This figure shows the eigenvalues of a single 1000 by 1000 iid random matrix with atom distribution x defined by $\mathbf{P}(x = 1) = \mathbf{P}(x = -1) = 1/2$ which was perturbed by adding the random matrix μB_n as defined after (2), where $\mu = 2$ and $p = 1/4$. (Figure by Phillip Wood.)

1.1. **Zero-sum matrices.** Next, we consider a different low-rank perturbation of an iid matrix model, in which the row sums of the matrix are forced to equal zero. Introduce the orthogonal projection matrix

$$P_n = 1 - \phi_n \phi_n^* = (\delta_{ij} - \frac{1}{n})_{1 \leq i, j \leq n};$$

thus P_n is the orthogonal projection onto the hyperplane $\phi_n^\perp = \{(x_1, \dots, x_n) \in \mathbb{C}^n : x_1 + \dots + x_n = 0\}$.

Our first result is that the presence of this projection does not affect the circular law, nor does it create any outliers:

Theorem 1.13. *Let X_n be an iid random matrix. Then $\mu \frac{1}{\sqrt{n}} X_n P_n$ converges both in probability and almost surely to the circular measure μ_c , and almost surely one has $\rho(\frac{1}{\sqrt{n}} X_n P_n) = 1 + o(1)$.*

We prove this theorem in Section 6, using the machinery from [44]. The main difficulty is to ensure that the least singular value of $\frac{1}{\sqrt{n}} X_n P_n - z$ is well controlled for any fixed z , but this can be achieved dropping one dimension (and freezing some of the entries of X_n) to eliminate the role of P_n (at the cost of replacing the deterministic matrix $z = zI$ by a more complicated matrix). This result is somewhat analogous to the result of [17] establishing the circular law for random Markov matrices (under an additional bounded density hypothesis on the atom distribution), although the situation is more complicated in that case because a slightly nonlinear transformation is required in order to convert a iid random matrix to a Markov matrix, in contrast to the simple linear transformation $X_n \mapsto X_n P_n$ required to make the rows sum to zero.

Suppose that ψ_n is a vector orthogonal to ϕ_n , then $\psi_n^* = \psi_n^* P_n$ and $P_n \phi_n = 0$. Then from two applications of (1) we have

$$\begin{aligned} \det \left(1 + z \left(\frac{1}{\sqrt{n}} X_n P_n + \phi_n \psi_n^* \right) \right) &= \det \left(1 + z P_n \left(\frac{1}{\sqrt{n}} X_n P_n + \phi_n \psi_n^* \right) \right) \\ &= \det \left(1 + z P_n \frac{1}{\sqrt{n}} X_n P_n \right) \\ &= \det \left(1 + z \frac{1}{\sqrt{n}} X_n P_n \right). \end{aligned}$$

We thus see that $\frac{1}{\sqrt{n}} X_n P_n$ and $\frac{1}{\sqrt{n}} X_n P_n + \phi_n \psi_n^*$ have the same characteristic polynomial, and thus the same ESD and the same spectral radius. (One can also establish these facts directly without much difficulty, as was done in [39].) We thus obtain

Corollary 1.14. *Let X_n be an iid random matrix, and for each n , let ψ_n be a (possibly random, and X_n -dependent) vector which is orthogonal to ϕ_n . Then $\mu_{\frac{1}{\sqrt{n}} X_n P_n + \phi_n \psi_n^*}$ converges both in probability and almost surely to the circular measure μ_c , and almost surely one has $\rho(\frac{1}{\sqrt{n}} X_n P_n + \phi_n \psi_n^*) = 1 + o(1)$.*

In particular, no outliers are created no matter how large ψ_n is (or how aligned it is with ϕ_n). These results were established in [39] in the gaussian case. This is in sharp contrast to the situation in Theorem 1.11 for the model (3), which is similar to the matrix $\frac{1}{\sqrt{n}} X_n P_n + \phi_n \psi_n^*$, but without the P_n projection.

Remark 1.15. Our results here are only effective in the region where the spectral parameter z has magnitude larger than that of the spectral radius. It would be of interest to determine what happens in models where the eigenvalue law of the base matrix $\frac{1}{\sqrt{n}} X_n$ is not governed by a circular law, but by another law whose support does not occupy the entire disk given by the spectral radius (e.g. matrices whose ESD is concentrated in an annulus). In the covariance matrix case, results in this direction appear in [5], [7].

1.2. Acknowledgments. This work was conducted during a workshop at the American Institute of Mathematics. We thank Larry Abbott for raising these questions. After posing this question at the workshop, Alice Guionnet provided the key insight, namely to reduce matters to studying coefficients of the resolvent of $\frac{1}{\sqrt{n}} X_n$, while Percy Deift emphasised the significance of the identity (1) to questions of this type (and indeed, this identity is crucial in order to efficiently handle the higher rank case $k > 1$). The author also thanks Sasha Soshnikov and Phillip Wood for useful discussions, Florent Benaych-Georges, Djalil Chafai and Raj Rao for references, and Phillip Wood for corrections. We are also indebted to Phillip Wood for supplying the figures for this paper. Finally, we thank the anonymous referees for many helpful comments, corrections, and references.

1.3. Asymptotic notation. Throughout this paper, n is an asymptotic parameter going to infinity. We use $o(1)$ to denote any quantity that is bounded in magnitude by an expression $c(n)$ that goes to zero as $n \rightarrow \infty$, keeping all other parameters independent of n (e.g. x, ε, k, z) fixed. Similarly, we use $X = O(Y)$ or $X \ll Y$ to denote the estimate $|X| \leq CY$ where the implied constant C is independent of n but may depend on parameters independent of n .

2. SMALL LOW RANK PERTURBATION

We now prove Theorem 1.7. Fix $\varepsilon > 0$ and x, C_n as in that theorem. By hypothesis, C_n has rank at most k for some $k = O(1)$ independent of n , and an operator norm of $O(1)$. By the singular value decomposition, we can write $C_n = A_n B_n$ for some $n \times k$ and $k \times n$ matrices A_n, B_n , both of operator norm $O(1)$.

We have the following description⁴ of the eigenvalues of $\frac{1}{\sqrt{n}}X_n + C_n$ in terms of a $k \times k$ determinant:

Lemma 2.1 (Eigenvalue criterion). *Let z be a complex number that is not an eigenvalue of $\frac{1}{\sqrt{n}}X_n$. Then z is an eigenvalue of $\frac{1}{\sqrt{n}}X_n + C_n$ if and only if*

$$(5) \quad \det(1 + B_n(\frac{1}{\sqrt{n}}X_n - z)^{-1}A_n) = 0.$$

Proof. Clearly, z is an eigenvalue of $\frac{1}{\sqrt{n}}X_n + C_n$ if and only if

$$\det(\frac{1}{\sqrt{n}}X_n + C_n - z) = 0.$$

By hypothesis, $\frac{1}{\sqrt{n}}X_n - z$ is invertible and $C_n = A_n B_n$, so we may rewrite this equation as

$$\det(1 + (\frac{1}{\sqrt{n}}X_n - z)^{-1}A_n B_n) = 0.$$

The claim now follows from (1). □

Remark 2.2. The above argument in fact shows that

$$\begin{aligned} \det(1 + B_n(\frac{1}{\sqrt{n}}X_n - z)^{-1}A_n) &= \frac{\det(\frac{1}{\sqrt{n}}X_n + C_n - z)}{\det(\frac{1}{\sqrt{n}}X_n - z)} \\ &= \frac{\prod_{j=1}^n (\lambda_j(\frac{1}{\sqrt{n}}X_n + C_n) - z)}{\prod_{j=1}^n (\lambda_j(\frac{1}{\sqrt{n}}X_n) - z)} \end{aligned}$$

whenever the denominator is nonzero. We are indebted to Alice Guionnet for suggesting the use of this type of criterion, versions of which also appear in [3], [8], [18], [13], [14], [15]. For instance, the $k = 1$ case of this criterion appears explicitly in [13]. In [3] the expression in this identity (which is stated there in the case of symmetric matrices) is referred to as the *modified Weinstein determinant*. (We thank Raj Rao for this reference.)

Introduce the functions

$$\begin{aligned} f(z) &:= \det(1 + B_n(\frac{1}{\sqrt{n}}X_n - z)^{-1}A_n) \\ g(z) &:= \det(1 + B_n(-z)^{-1}A_n). \end{aligned}$$

These are both meromorphic functions that are asymptotically equal to 1 at infinity, with g being a rational function of degree at most k with bounded coefficients. Lemma 2.1 tells us that outside of the spectrum of $\frac{1}{\sqrt{n}}X_n$, the zeroes of $f(z)$ agree with the eigenvalues of $\frac{1}{\sqrt{n}}X_n + C_n$. An inspection of the argument also reveals that the multiplicity of a given such eigenvalue is equal to the degree of the corresponding zero of f . Similarly, replacing $\frac{1}{\sqrt{n}}X_n$ by the zero matrix in Lemma 2.1, we see that outside of the origin, the zeroes of g are precisely the eigenvalues of C_n (counting multiplicity). Indeed, from (1) one has

$$g(z) = \prod_{i=1}^k \left(1 - \frac{\lambda_i(C_n)}{z}\right),$$

where $\lambda_1(C_n), \dots, \lambda_k(C_n)$ are the k non-trivial eigenvalues of C_n (some of which may be zero), including of course the j eigenvalues $\frac{1}{\sqrt{n}}X_n$ of magnitude at least $1 + 3\varepsilon$.

⁴We are indebted to Alice Guionnet for proposing the $k = 1$ case of this formula, as well as the basic strategy of proof used in this paper, and Percy Deift for emphasising the importance of the identity (1).

By Theorem 1.4, we see that almost surely, the spectrum of $\frac{1}{\sqrt{n}}X_n$ is contained in the disk $\{z \in \mathbb{C} : |z| \leq 1 + \varepsilon\}$ for sufficiently large n . In view of Rouché's theorem (or the argument principle), together with the fact that the coefficients of g are bounded, it then suffices to show that the quantity

$$\sup_{|z| \geq 1+2\varepsilon} |f(z) - g(z)|$$

converges almost surely to zero. Since k is fixed and B_n, A_n are bounded in operator norm, it suffices to show that

$$\sup_{|z| \geq 1+2\varepsilon} \|B_n((\frac{1}{\sqrt{n}}X_n - z)^{-1} - (-z)^{-1})A_n\|_{op}$$

converges almost surely to zero.

By Theorem 1.4, we almost surely have

$$\left\| \left(\frac{1}{\sqrt{n}}X_n \right)^m \right\|_{op} = m + 1 + o(1)$$

for each $m \geq 1$. In particular, there exists $m_0 > 0$ such that

$$\left\| \left(\frac{1}{\sqrt{n}}X_n \right)^{m_0} \right\|_{op} \leq (1 + \varepsilon)^{m_0}$$

for sufficiently large n , which also implies that one has

$$\left\| \left(\frac{1}{\sqrt{n}}X_n \right)^m \right\|_{op} \leq K(1 + \varepsilon)^m$$

for all $n, m \geq 1$ and some almost surely finite random variable K . This ensures that the Neumann series

$$\left(\frac{1}{\sqrt{n}}X_n - z \right)^{-1} - (-z)^{-1} = - \sum_{m=1}^{\infty} z^{-m-1} \left(\frac{1}{\sqrt{n}}X_n \right)^m$$

is absolutely convergent in the operator norm, uniformly in both n and z , when n is sufficiently large and $|z| \geq 1 + 2\varepsilon$. By the dominated convergence theorem, it thus suffices to show that the $k \times k$ matrix

$$B_n \left(\frac{1}{\sqrt{n}}X_n \right)^m A_n$$

converges almost surely to zero for each fixed $m \geq 1$. Breaking B_n and A_n into components, it suffices to show the following claim:

Lemma 2.3 (Coefficient bound). *Let X_n be an iid random matrix whose atom distribution has finite fourth moment. Then*

$$(6) \quad \left\langle \left(\frac{1}{\sqrt{n}}X_n \right)^m u_n, v_n \right\rangle = o(1)$$

almost surely for each fixed $m \geq 1$ and any fixed (deterministic) sequence of unit vectors $u_n, v_n \in \mathbb{C}^n$.

We now prove the lemma.

The atom distribution x is currently only assumed to have finite fourth moment. However, a standard truncation argument (using the results of [9] to control the contribution of the tail of x) shows that we may almost surely approximate $\frac{1}{\sqrt{n}}X_n$ to arbitrary accuracy in operator norm by an iid matrix in which the atom distribution is in fact bounded. As such, it will suffice to prove the lemma under the additional assumption that x is bounded. In particular, all moments of x are now finite: $\mathbf{E}|x|^j = O(1)$ for all fixed j .

By diagonalising the covariance matrix of $\text{Re}(x)$ and $\text{Im}(x)$, we may assume (after a phase rotation) that $\text{Re}(x)$ and $\text{Im}(z)$ have zero covariance, and have variances σ^2 and $1 - \sigma^2$ respectively for some $0 \leq \sigma^2 \leq 1$. Next, by splitting u_n, v_n into real and imaginary parts and renormalising, we may assume without loss of generality that u_n, v_n have real coefficients.

We now use the second moment method. It will suffice to show that

$$\mathbf{E}|\langle (\frac{1}{\sqrt{n}}X_n)^m u_n, v_n \rangle|^2 = o(1).$$

We first deal with the model case when x is normally distributed with the the normal distribution $N(0, \sigma^2)_{\mathbb{R}} + iN(0, 1 - \sigma^2)_{\mathbb{R}}$ (with the real and imaginary parts independent). As is well known, in this case the ensemble G_n is invariant under left and right multiplication by orthogonal matrices. This implies that the expression $\mathbf{E}|\langle (\frac{1}{\sqrt{n}}G_n)^m u_n, v_n \rangle|^2$ is in fact independent of the choice of the real unit vectors u_n, v_n . Letting u_n, v_n range over the standard basis e_1, \dots, e_n and averaging, we conclude that

$$\begin{aligned} \mathbf{E}|\langle (\frac{1}{\sqrt{n}}G_n)^m u_n, v_n \rangle|^2 &= \frac{1}{n^2} \mathbf{E} \text{trace}(\frac{1}{\sqrt{n}}G_n)^m ((\frac{1}{\sqrt{n}}G_n)^m)^* \\ &\leq \frac{1}{n} \mathbf{E} \|\frac{1}{\sqrt{n}}G_n\|_{op}^m. \end{aligned}$$

By Theorem 1.4, we thus obtain a bound of $O(1/n) = o(1)$ in this case.

To handle the non-gaussian case, it thus suffices to show that

$$\mathbf{E}|\langle (\frac{1}{\sqrt{n}}X_n)^m u_n, v_n \rangle|^2 - \mathbf{E}|\langle (\frac{1}{\sqrt{n}}G_n)^m u_n, v_n \rangle|^2 = o(1).$$

We expand the left-hand side as

$$\begin{aligned} n^{-m} \sum_{a_0, \dots, a_m, b_0, \dots, b_m \in \{1, \dots, n\}} u_{n,a_0} \overline{v_{n,a_m}} \overline{u_{n,b_0}} v_{n,b_m} \\ \left(\mathbf{E} \prod_{j=1}^m x_{a_j, a_{j-1}} \overline{x_{b_j, b_{j-1}}} - g_{a_j, a_{j-1}} \overline{g_{b_j, b_{j-1}}} \right), \end{aligned}$$

where $u_{n,i}, v_{n,i}$ are the coordinates of the unit vectors u_n, v_n , and g_{ij} are iid copies of the complex normal distribution $N(0, \sigma^2)_{\mathbb{R}} + iN(0, 1 - \sigma^2)_{\mathbb{R}}$.

Consider the collection of $2m$ ordered pairs $(a_{j-1}, a_j), (b_{j-1}, b_j)$ with $j = 1, \dots, m$. From the iid nature of X , and the fact that the random variables x and g match up to second order, we see that each summand in the above expression vanishes unless each ordered pair appears with multiplicity at least two, and at least one pair appears at least three times; in particular, there are at most $m - 1$ distinct ordered pairs. As x, g have all moments finite, we may thus bound the above expression in magnitude by

$$O(n^{-m} \sum_{*} |u_{n,a_0}| |v_{n,a_m}| |u_{n,b_0}| |v_{n,b_m}|)$$

where the sum ranges over all tuples $a_0, \dots, a_m, b_0, \dots, b_m$ which generate at most $m - 1$ distinct ordered pairs.

By the arithmetic mean-geometric mean inequality, we may bound

$$|u_{n,a_0}| |v_{n,a_m}| |u_{n,b_0}| |v_{n,b_m}| \ll |u_{n,a_0}|^2 |u_{n,b_0}|^2 + |v_{n,a_m}|^2 |u_{n,b_0}|^2 + |u_{n,a_0}|^2 |v_{n,b_m}|^2 + |v_{n,a_m}|^2 |v_{n,b_m}|^2,$$

thus bounding the previous expression by the sum of four terms, of which

$$(7) \quad O(n^{-m} \sum_{*} |u_{n,a_0}|^2 |v_{n,b_m}|^2)$$

is typical. We will just show that the stated term (7) is $o(1)$; the other three terms are treated by a similar argument.

A given tuple in the sum \sum_{*} has at most $m - 1$ distinct ordered pairs, which then gives rise to an unordered, looped graph on $\{1, \dots, n\}$ that is the union of two paths of length m , and thus either has one or two connected components (excluding vertices with no edges). Suppose first that the graph has one

connected component. Then, after fixing a_0 , there are only $O(n^{m-1})$ possible ways to choose the remaining components of the tuple; bounding v_{n,b_m} crudely by $O(1)$, we see that the contribution of this case to (7) is

$$O\left(n^{-m} \sum_{a_0=1}^n |u_{n,a_0}|^2 n^{m-1}\right) = O(n^{-1}) = o(1)$$

which is acceptable.

Now suppose that there are two connected components, then a_0, \dots, a_m and b_0, \dots, b_m occupy distinct components. After fixing both a_0 and b_m , there are at most $O(n^{m-1})$ ways to choose the remaining components of the tuple; this gives a contribution of

$$O(n^{-m} \sum_{a_0=1}^n \sum_{b_m=1}^n |u_{n,a_0}|^2 |v_{n,b_m}|^2 n^{m-1}) = O(n^{-1}) = o(1)$$

which is again acceptable. This completes the proof of Lemma 2.3 and hence Theorem 1.7.

Remark 2.4. One could also use concentration of measure inequalities (see e.g. [30]) to establish Lemma 2.3, but the moment methods used above will be useful when we establish a central limit theorem variant of Lemma 2.3 in Section 4 below.

3. LARGE MEAN

Now we prove Theorem 1.9. It suffices to show that for any fixed $\varepsilon > 0$, almost surely one has exactly one eigenvalue of $\frac{1}{\sqrt{n}}X_n + \mu\sqrt{n}\phi_n\phi_n^*$ outside of the disk $\{z : |z| \leq 1 + 2\varepsilon\}$, with this eigenvalue occurring within $O(\varepsilon)$ of $\mu\sqrt{n}$.

Fix ε . By Theorem 1.4, almost surely there are no eigenvalues of $\frac{1}{\sqrt{n}}X_n$ outside of the disk. By Lemma 2.1, the eigenvalues of $\frac{1}{\sqrt{n}}X_n + \mu\sqrt{n}\phi_n\phi_n^*$ outside this disk are then precisely the solutions (counting multiplicity) to the equation $f(z) = 0$, where f is the meromorphic function

$$f(z) := 1 + \mu\sqrt{n} \left\langle \left(\frac{1}{\sqrt{n}}X_n - z\right)^{-1} \phi_n, \phi_n \right\rangle.$$

By Neumann series, we may expand

$$f(z) = g(z) - \mu\sqrt{n} \sum_{m=1}^{\infty} z^{-m-1} \left\langle \left(\frac{1}{\sqrt{n}}X_n\right)^m \phi_n, \phi_n \right\rangle$$

where

$$(8) \quad g(z) := 1 - \mu\sqrt{n}/z.$$

From Theorem 1.4, we almost surely have

$$\left\| \left(\frac{1}{\sqrt{n}}X_n\right) \right\|_{op} \leq 3$$

(say) and

$$\left\| \left(\frac{1}{\sqrt{n}}X_n\right)^{m_0} \right\|_{op} \leq (1 + \varepsilon)^{m_0}$$

for some fixed integer m_0 depending only on ε , and all sufficiently large n ; this gives us the truncated Taylor expansion

$$f(z) = g(z) - \mu\sqrt{n} \sum_{m=1}^M z^{-m-1} \left\langle \left(\frac{1}{\sqrt{n}}X_n\right)^m \phi_n, \phi_n \right\rangle + O\left(\frac{(1 + \varepsilon)^M}{|z|^{M+2}} \sqrt{n}\right)$$

for any $M \geq 1$ and $|z| \geq 1 + 2\varepsilon$, where the implied constant is allowed to depend on ε, μ but not on M . Applying Lemma 2.3, we almost surely obtain

$$f(z) = g(z) + o\left(\frac{\sqrt{n}}{|z|^2}\right) + O\left(\frac{(1+\varepsilon)^M}{|z|^{M+2}}\sqrt{n}\right)$$

for any fixed M uniformly for all $|z| \geq 1 + 2\varepsilon$, and thus (by letting M go slowly to infinity) we almost surely have

$$f(z) = g(z) + o\left(\frac{\sqrt{n}}{|z|^2}\right)$$

uniformly for all $|z| \geq 1 + 2\varepsilon$. From this and (8) we see that f has no zeroes in this region except within a $o(1)$ neighbourhood of $\mu\sqrt{n}$, and from Rouché's theorem we see that there is exactly one zero of that latter type when n is sufficiently large, and the claim follows.

Remark 3.1. The eigenvector corresponding to the exceptional zero can be explicitly described, by observing the identity

$$\left(\frac{1}{\sqrt{n}}X_n + \mu\sqrt{n}\phi_n\phi_n^* - z\right)\left(1 - \frac{1}{z\sqrt{n}}X_n\right)^{-1}\phi_n = \frac{-f(z)}{z}\phi_n$$

for all non-zero z outside of the spectrum of $\frac{1}{\sqrt{n}}X_n$. In particular, if $z = \mu\sqrt{n} + o(1)$ is the outlier eigenvalue, the $(1 - \frac{1}{z\sqrt{n}}X_n)^{-1}\phi_n$ is the corresponding eigenvector. From Theorem 1.4 and Neumann series, we see that almost surely, this eigenvector lies within $O(1/\sqrt{n})$ of ϕ_n in ℓ^2 norm, and a more accurate description of this eigenvector can be given by expanding out the Neumann series further.

4. A CENTRAL LIMIT THEOREM

In Lemma 2.3, we showed that the coefficients $\langle (\frac{1}{\sqrt{n}}X_n)^m u_n, v_n \rangle$ decayed almost surely to zero. Now we prove a more refined statement on the rate of decay, which is to Lemma 2.3 as the central limit theorem is to the (strong) law of large numbers. This result will be needed to prove Theorem 1.11. For simplicity we consider only real-valued matrices; there is a complex analogue when the real and imaginary parts of x have the same covariance matrix as the complex Gaussian $N(0, 1)_{\mathbb{C}}$, but we will not state it here.

Proposition 4.1 (Central limit theorem). *Let X_n be an iid random matrix whose atom distribution is real-valued, has mean zero and has all moments finite. Let $u_n, v_n \in \mathbb{R}^n$ be a (deterministic) sequence of unit vectors whose coefficients $u_{n,i}, v_{n,i}$ are asymptotically delocalised in the sense that*

$$(9) \quad \sup_{1 \leq i \leq n} |u_{n,i}|, \sup_{1 \leq i \leq n} |v_{n,i}| = O(1/\sqrt{n}).$$

Then for any fixed $m \geq 1$, the m random variables

$$(10) \quad Z_j := \sqrt{n} \left\langle \left(\frac{1}{\sqrt{n}}X_n\right)^j u_n, v_n \right\rangle$$

for $j = 1, \dots, m$ converge jointly in distribution to the law of m independent copies of the real gaussian $N(0, 1)_{\mathbb{R}}$.

We now prove this proposition. By (the multidimensional version of) Carleman's theorem (see e.g. [6]), it suffices to prove the moment bounds

$$(11) \quad \mathbf{E} \prod_{j=1}^m Z_j^{r_j} = \mathbf{E} \prod_{j=1}^m G_j^{r_j} + o(1)$$

for any natural numbers r_1, \dots, r_m , where G_1, \dots, G_m are iid copies of the real gaussian $N(0, 1)_{\mathbb{R}}$.

Fix m, r_1, \dots, r_m ; we allow all implied constants to depend on these quantities. The left-hand side can be expanded as

$$(12) \quad n^{-A} \sum_{*} \mathbf{E} \prod_{j=1}^m \left(\prod_{l=1}^{r_j} U_{a_{j,l,0}} V_{a_{j,l,j}} \right) \left(\prod_{l=1}^{r_j} \prod_{i=1}^j x_{a_{j,l,i}, a_{j,l,i-1}} \right)$$

where $U_i := \sqrt{n}u_{n,i}$, $V_i := \sqrt{n}v_{n,i}$,

$$A := \frac{1}{2} \sum_{j=1}^m r_j(j+1)$$

and $*$ ranges over all tuples of indices $a_{j,l,i} \in \{1, \dots, n\}$ with $1 \leq j \leq m$, $1 \leq l \leq r_j$, $0 \leq i \leq j$.

From (9) and the bounded moment hypotheses, we see that each summand (12) is of size $O(1)$, and so we may freely ignore up to $o(n^A)$ summands whenever desired.

For each tuple in the sum \sum_{*} , we consider the $\sum_{j=1}^m j r_j$ ordered pairs $(a_{j,l,i-1}, a_{j,l,i})$ with $1 \leq j \leq m$, $1 \leq l \leq r_j$, $1 \leq i \leq j$. By the iid and mean zero nature of the $x_{i,j}$, we see that this sum vanishes unless each ordered pair appears with multiplicity at least two. In particular each of the indices $a_{j,l,i}$ with $0 \leq i \leq j$ must occur with multiplicity at least 2, leading to at most A distinct indices. If there are any fewer than A distinct indices, then the contribution of this case to (12) is $o(n^A)$, so we may assume that there are exactly A distinct indices, which implies that each index $a_{j,l,i}$ occurs with multiplicity exactly two. This implies that for $1 \leq i \leq j$, each of the $a_{j,l,i-1}$ arise exactly twice as the initial vertex of an ordered pair, and each of the $a_{j,l,i}$ arise exactly twice as the final vertex of an ordered pair. Furthermore, the indices $a_{j,l,0}$ must be distinct from the indices $a_{j,l,i}$ with $0 < i \leq j$, and similarly the $a_{j,l,j}$ must be distinct from the indices $a_{j,l,i}$ with $i < j$, as otherwise the fact that each ordered pair appears at least twice will lead to a multiplicity of at least three at the repeated index. This implies that the $\sum_j r_j$ paths $(a_{j,l,0}, \dots, a_{j,l,j})$ are simple paths, which each occur with multiplicity two but are otherwise disjoint. The total contribution of this case to (12) is then

$$n^{-A} \sum_{**} \prod_{j=1}^m \left(\prod_{l=1}^{r_j} U_{a_{j,l,0}} V_{a_{j,l,j}} \right),$$

where the sum \sum_{**} is over collections of simple paths $(a_{j,l,0}, \dots, a_{j,l,j})$ in $\{1, \dots, n\}$ which occur with multiplicity two but are otherwise disjoint; here we use the fact that each ordered pair appears exactly twice, and that x has unit variance.

In order for all paths to appear with multiplicity two, each of the r_j must be even. This already gives (11) when at least one of the r_j is odd, so we now assume that the r_j are all even. There are $\prod_{j=1}^m \frac{r_j!}{2^{r_j/2} (r_j/2)!}$ different ways in which the paths can be matched up to multiplicity two. Once one fixes such a matching, there are $R := \frac{1}{2} \sum_{j=1}^m r_j$ initial vertices b_1, \dots, b_R of paths, and R final vertices c_1, \dots, c_R , all distinct from each other; if one fixes these vertices, then one has $(1 + o(1)) n^{\sum_{j=1}^m (j-1)r_j/2} = (1 + o(1)) n^{A-2R}$ ways to choose the remaining paths. This gives a total contribution to (12) of

$$(1 + o(1)) n^{-2R} \prod_{j=1}^m \frac{r_j!}{2^{r_j/2} (r_j/2)!} \sum_{b_1, \dots, b_R, c_1, \dots, c_R \text{ distinct}} \prod_{r=1}^R U_{b_r}^2 V_{c_r}^2.$$

We add back in the contributions in which some of the $b_1, \dots, b_R, c_1, \dots, c_R$; this only affects the sum by $o(n^{2R})$, which is acceptable. We are left with

$$(1 + o(1)) n^{-2R} \prod_{j=1}^m \frac{r_j!}{2^{r_j/2} (r_j/2)!} \sum_{b_1, \dots, b_R, c_1, \dots, c_R \in \{1, \dots, n\}} \prod_{r=1}^R U_{b_r}^2 V_{c_r}^2;$$

since the U_b and V_c square-sum to n , this simplifies to

$$(1 + o(1)) \prod_{j=1}^m \frac{r_j!}{2^{r_j/2} (r_j/2)!}$$

and (11) follows from the standard computation

$$\mathbf{E}G^r = \frac{r!}{2^{r/2}(r/2)!}$$

for r even.

5. LARGE NON-SELFADJOINT PERTURBATION

We now prove Theorem 1.11. It suffices to work in the exterior region $\Omega := \{z : |z| \geq 1 + 4\varepsilon\}$ for a fixed $\varepsilon > 0$. Henceforth all implied constants may depend on ε, p, μ .

5.1. Crude upper bound. We first show the first part of the theorem, namely that $\sup_n \mathbf{E}N_{4\varepsilon}^m < \infty$ for all $m \geq 1$. Fix m ; we allow all implied constants to depend on m . Our task is now to show that $\mathbf{E}N_{4\varepsilon}^m = O(1)$ for all sufficiently large n .

From Theorem 1.4 we know that the spectral radius of $\frac{1}{\sqrt{n}}M_n$ is at most $1 + \varepsilon$ with overwhelming probability. Using the trivial bound $N_{4\varepsilon} \leq n$, we see that the tail event when the spectral radius exceeds $1 + \varepsilon$ thus gives a negligible contribution to $\mathbf{E}N_{4\varepsilon}^m$ and will thus be ignored.

Conditioning on the event that the spectral radius is at most $1 + \varepsilon$, we may then apply Lemma 2.1 to conclude that the eigenvalues of A_n in Ω are precisely the zeroes in Ω of the random analytic function

$$(13) \quad f(z) := 1 + \mu\sqrt{n} \left\langle \left(\frac{1}{\sqrt{n}}X_n - zI \right)^{-1} \phi_n, \psi_n \right\rangle.$$

In particular, $N_{4\varepsilon}$ is the number of zeroes of f in Ω .

The function $f(1/z)$ is analytic in the disk $\{w \in \mathbb{C} : |w| < \frac{1}{1+\varepsilon}\}$ and equals 1 at the origin, with $N_{4\varepsilon}$ zeroes in the region $\{w \in \mathbb{C} : |w| < \frac{1}{1+4\varepsilon}\}$. Applying Jensen's formula to this function, we conclude the upper bound

$$N_{4\varepsilon} \ll \int_{|z|=r} \log_+ \frac{1}{|f(z)|} |dz|$$

for any radius r between $1 + 2\varepsilon$ and $1 + 3\varepsilon$ (note that we allow implied constants to depend on ε), where $\log_+ x := \max(\log x, 0)$ and $|dz|$ is arclength measure. Averaging, we conclude that

$$(14) \quad N_{4\varepsilon} \ll \int_{1+2\varepsilon \leq |z| \leq 1+3\varepsilon} \log_+ \frac{1}{|f(z)|} d^2z.$$

It will thus suffice to establish the bound

$$(15) \quad \mathbf{E} \int_K \log_+^m \frac{1}{|f(z)|} \ll 1$$

for any fixed $m \geq 1$ and any compact subset K of the annulus $|z| > 1 + \varepsilon$ (allowing implied constants to depend on m, K of course).

We now pause to regularise the logarithm slightly, as this will come in handy later. By Remark 2.2, the function $\log \frac{1}{|f(z)|}$ is a linear combination of $O(n)$ terms of the form $\log |z - z_0|$ for various complex numbers z_0 . As such we have a crude upper bound of the form

$$(16) \quad \int_K \log_+^{m+1} \frac{1}{|f(z)|} d^2z \ll n^{m+1}$$

thanks to the triangle inequality. From this, we have

$$\int_K 1_{\log_+ \frac{1}{|f(z)|} \geq n^{m+1}} \log_+^m \frac{1}{|f(z)|} d^2z \ll 1,$$

and so it suffices to show that

$$\mathbf{E} \int_K \min(\log_+^m \frac{1}{|f(z)|}, n^{m+1}) d^2 z \ll 1.$$

By the Fubini-Tonelli theorem, it thus suffices to show that

$$\mathbf{E} \min \left(\log_+ \frac{1}{|f(z)|}, n^{m+1} \right)^m = O(1)$$

uniformly for all z in K .

It will then suffice to show the following lower tail estimates on f :

Lemma 5.1 (Lower tail estimates). *Let z be in a compact subset K of $\{z \in \mathbb{C} : |z| > 1 + \varepsilon\}$; we allow implied constants to depend on K .*

(i) *For every $A > 0$ there exists $B > 0$ (not depending on n) such that*

$$\mathbf{P}(|f(z)| \leq n^{-B}) \ll n^{-A}.$$

(ii) *For every $\delta > 0$ one has*

$$\mathbf{P}(|f(z)| \leq \delta) \ll \delta + n^{-c}$$

for some absolute constant $c > 0$.

Indeed, item (i) (with $A = 2m$) allows one to reduce to the case when $\log_+ \frac{1}{|f(z)|} = O(\log n)$ (it is here that we take advantage of our previous regularisation of the logarithm), and then (ii) and a dyadic decomposition gives the claim.

Proof. We begin with (i). From (3) and (13) we have the identity

$$(A_n - zI) \left(\frac{1}{\sqrt{n}} X_n - zI \right)^{-1} \phi_n = f(z) \phi_n$$

and hence

$$|f(z)| \geq \frac{\sigma_n(A_n - zI)}{\left\| \frac{1}{\sqrt{n}} X_n - zI \right\|_{op}},$$

where $\sigma_n(A_n - zI)$ is the least singular value of $A_n - zI$. By Theorem 1.4, we have $\left\| \frac{1}{\sqrt{n}} X_n - zI \right\|_{op} = O(1)$ with overwhelming probability. The claim now follows from the least singular value bounds in [44, Lemma 4.1], [42, Theorem 2.1], or [43, Theorem 4.1], since we may express $A_n - zI$ as the sum of the normalised iid random matrix $\frac{1}{\sqrt{n}} X_n$ and the deterministic matrix $\mu \sqrt{n} \phi_n \psi_n^* - zI$, which has polynomial size.

Now we prove (ii). Let v be the vector $v := \left(\frac{1}{\sqrt{n}} X_n - zI \right)^{-1} \phi_n$. It will suffice to show that

$$\mathbf{P}(\sqrt{n} \langle v, \psi_n \rangle \in I) \ll |I| + n^{-c}$$

for every interval I . From Theorem 1.4 and Neumann series, we see that with overwhelming probability,

$$(17) \quad \left\| \left(\frac{1}{\sqrt{n}} X_n - zI \right)^{-1} \right\|_{op}, \left\| \frac{1}{\sqrt{n}} X_n - zI \right\|_{op} = O(1)$$

and so

$$1 \ll \|v\| \ll 1.$$

Let $\delta > 0$ be a small quantity (independent of n) to be chosen later. Call v *delocalised* if we have $|v_i| \geq \delta/\sqrt{n}$ for at least δn of the indices $i = 1, \dots, n$. As the coefficients of ψ_n are iid (and are independent of v), we see from the Berry-Esséen theorem (see e.g. [20, Chapter XVI]) that the contribution of the delocalised v are acceptable. It thus suffices to show that v is delocalised with probability $1 - O(n^{-c})$ for some $c > 0$.

We will use an epsilon-net argument (cf. [31], [40]). If v is not delocalised, then v lies within $O(\delta)$ in ℓ^2 norm of a vector w of comparable to 1 that is *sparse* in the sense that it is supported on at most δn indices,

simply by restricting v to those indices i for which $|v_i| \geq \delta/\sqrt{n}$, and truncating all other coefficients to zero. From (17) and the triangle inequality, we thus have

$$\|(\frac{1}{\sqrt{n}}X_n - zI)w - \phi_n\| \ll \delta.$$

The number of possible supports for w is at most $\exp(O(\delta \log \frac{1}{\delta} \times n))$, thanks to Stirling's formula. Once one fixes the supports, one can cover the range of w by $O((1/\delta))^{\delta n}$ balls in ℓ^2 norm of radius δ . Thus, by moving w by at most $O(\delta)$ if necessary, we may assume that w lies in a net Σ of cardinality

$$(18) \quad |\Sigma| \ll \exp(O(\delta \log \frac{1}{\delta} \times n)).$$

For each fixed w , the expression $\sqrt{n}\|(\frac{1}{\sqrt{n}}X_n - zI)w - \phi_n\|$ is a convex, 1-Lipschitz function of X_n as measured using the Frobenius norm $\|A\|_F := (\text{trace}(A^*A))^{1/2}$. Applying⁵ either the Talagrand concentration inequality (if x is bounded) or the Lévy concentration inequality (if x is Gaussian), we conclude that

$$\mathbf{P}\left(\left|\sqrt{n}\|(\frac{1}{\sqrt{n}}X_n - zI)w - \phi_n\| - M\right| \geq \lambda\right) \ll e^{-c\lambda^2}$$

for some absolute constant $c > 0$, where M is the median value of $\sqrt{n}\|(\frac{1}{\sqrt{n}}X_n - zI)w - \phi_n\|$. To compute this median, we first compute the second moment

$$\mathbf{E}\left(\sqrt{n}\|(\frac{1}{\sqrt{n}}X_n - zI)w - \phi_n\|\right)^2.$$

Expanding this out and using the fact that X_n is an iid random matrix, this simplifies to

$$n(\|zw + \phi_n\|^2 + \|w\|^2).$$

In particular, this expression is comparable to n . From the concentration inequality, we conclude that

$$\mathbf{P}\left(\|(\frac{1}{\sqrt{n}}X_n - zI)w - \phi_n\| \ll \delta\right) \ll \exp(-cn)$$

if δ is small enough. Summing up over all $w \in \Sigma$ using (18) and the union bound, we obtain the claim. \square

This concludes the proof of part (i) of Theorem 1.11.

5.2. Correlation functions. Now we prove part (ii) of Theorem 1.11. From part (i), we have the bound

$$(19) \quad \int_{\Omega^k} \rho_n^{(k)}(z_1, \dots, z_k) d^2 z_1 \dots d^2 z_k = O(1)$$

for each fixed k . A similar (but simpler) argument also shows that

$$(20) \quad \int_{\Omega^k} \rho_\infty^{(k)}(z_1, \dots, z_k) d^2 z_1 \dots d^2 z_k = O(1)$$

(the point being that the Gaussian random Laurent series $g(z)$ has a Gaussian distribution at each z with an explicitly computable variance, so one can easily control the moments of $(\log^+ \frac{1}{|g(z)|})^m$). As a consequence of these bounds, we can control perturbations to the test functions F that are small in the uniform norm.

From Theorem 1.4 we know that the spectral radius of $\frac{1}{\sqrt{n}}M_n$ is at most $1+\varepsilon$ with overwhelming probability. The tail event when the spectral radius exceeds $1+\varepsilon$ is thus negligible for the purposes of computing the asymptotics of the correlation functions and will thus be ignored.

⁵For an extensive discussion of concentration inequalities, see [30]. Note that many other atom distributions also enjoy concentration inequalities, such as those distributions with the log-Sobolev property; we will not attempt to aim for maximal generality here.

Conditioning on the event that the spectral radius is at most $1 + \varepsilon$, we may then apply Lemma 2.1 to conclude that the eigenvalues of A_n in Ω are precisely the zeroes in Ω of the random analytic function

$$f(z) := 1 + \mu\sqrt{n} \left\langle \left(\frac{1}{\sqrt{n}} X_n - zI \right)^{-1} \phi_n, \psi_n \right\rangle.$$

Thus, up to errors of $o(1)$, the correlation function $\rho_{A_n}^{(k)}$ is equal on Ω to the correlation function of the zeroes of f (defined as in (4)).

As in previous sections, once the spectral radius is at most $1 + \varepsilon$, we can expand f as a convergent Neumann series

$$f(z) = 1 - \mu \sum_{j=0}^{\infty} \frac{f_{n,j}}{z^{j+1}}$$

where

$$f_{n,j} := \sqrt{n} \left\langle \left(\frac{1}{\sqrt{n}} X_n \right)^j \phi_n, \psi_n \right\rangle.$$

To control this expression properly, we will need to work instead with the truncated Neumann series

$$f(z) = 1 - \mu \sum_{j=0}^J \frac{f_{n,j}}{z^{j+1}} + \frac{\mu}{z^{J+1}} R_J(z)$$

for any $J \geq 1$, where the remainder R_J is given by the formula

$$R_J(z) := \sqrt{n} \left\langle \left(\frac{1}{\sqrt{n}} X_n \right)^{J+1} \left(\frac{1}{\sqrt{n}} X_n - zI \right)^{-1} \phi_n, \psi_n \right\rangle.$$

We now obtain a concentration bound on $R_J(z)$:

Lemma 5.2. *Let $A, J \geq 1$. If n is sufficiently large (depending on A, J, ε, k), then for each z with $|z| \geq 1 + 2\varepsilon$ and all $\lambda > 0$, one has*

$$\mathbf{P}(|R_J(z)| \geq \lambda J) \ll e^{-c\lambda^2} + n^{-A}$$

for some $c > 0$ depending only on ε, p, k .

Proof. Let m_0 be an integer such that $(m_0 + 1) < (1 + \varepsilon)^{m_0}$. By Theorem 1.4, we see that with overwhelming probability, we have

$$\left\| \left(\frac{1}{\sqrt{n}} X_n \right) \right\|_{op} \ll 1$$

and

$$\left\| \left(\frac{1}{\sqrt{n}} X_n \right)^{m_0} \right\|_{op} \leq (1 + \varepsilon)^{m_0}$$

and hence by Neumann series

$$\left\| \left(\frac{1}{\sqrt{n}} X_n - zI \right)^{-1} \right\|_{op} \ll 1$$

(recall that we allow implied constants to depend on ε). Henceforth we condition on the above event. By another application of Theorem 1.4, we see that with overwhelming probability, we have

$$\left\| \left(\frac{1}{\sqrt{n}} X_n \right)^{J+1} \right\|_{op} \ll J$$

and hence

$$\left\| \left(\frac{1}{\sqrt{n}} X_n \right)^{J+1} \left(\frac{1}{\sqrt{n}} X_n - zI \right)^{-1} \phi_n \right\| \ll J.$$

Note that the random vector ψ_n is independent of X_n . The claim then follows from the Azuma-Hoeffding inequality. \square

Meanwhile, we have the following law for the $f_{n,j}$:

Lemma 5.3. *Let $J \geq 1$. As $n \rightarrow \infty$, the random variables $f_{n,0}, \dots, f_{n,J}$ converge jointly in distribution to $J+1$ iid copies g_0, \dots, g_J of the real normal distribution $N(0, 1)_{\mathbb{R}}$.*

Proof. By the central limit theorem, $f_{n,0} = \sqrt{n}\langle \phi_n, \psi_n \rangle$ converges in distribution to $N(0, 1)_{\mathbb{R}}$. Now freeze ψ_n and thus $f_{n,0}$. By the law of large numbers, ψ_n has a norm of $1 + o(1)$ with probability $1 - o(1)$. Conditioning on this event, we see from Proposition 4.1 that $f_{n,1}, \dots, f_{n,J}$ converge in distribution to g_1, \dots, g_J . Integrating out the conditioning on ψ_n , we obtain the claim. \square

In view of this proposition and the Skorokhod representation theorem, we may thus find iid copies g_0, g_1, g_2, \dots of the real normal distribution $N(0, 1)_{\mathbb{R}}$ (depending on n) that are coupled to the $f_{n,j}$ in such a way that

$$(21) \quad \sup_{0 \leq j \leq J} |f_{n,j} - g_j| = o(1)$$

uniformly with probability $1 - o(1)$, for each $J \geq 1$. We thus have

$$f(z) = 1 - \mu \sum_{j=0}^J \frac{g_j}{z^{j+1}} + \frac{\mu}{z^{J+1}} R_J(z) + o(1)$$

uniformly with probability $1 - o(1)$, for any fixed J .

Next, we introduce the function

$$g(z) = 1 - \mu \sum_{j=0}^{\infty} \frac{g_j}{z^{j+1}}.$$

The correlation functions $\rho_{\infty}^{(k)}$ are the correlation functions of the zeroes of g . It thus suffices to show that the correlation functions of the zeroes of f converge in the vague topology to the correlation functions of the zeroes of g .

For any fixed z in Ω , the tail $\mu \sum_{j=J+1}^{\infty} \frac{g_j}{z^{j+1}}$ of $g(z)$ is Gaussian with mean zero and variance $O(|z|^{-2J-2})$. It thus obeys the same tail bound as Lemma 5.2 (indeed it obeys slightly better bounds). From this, Lemma 5.2, and (21) we conclude that

$$\mathbf{P}(|f(z) - g(z)| \geq \lambda J / |z|^{J+1}) \ll e^{-c' \lambda^2} + o(1)$$

for any $\lambda \geq 1$ and $J \geq 1$, where $c' > 0$ is an absolute constant and the decay rate $o(1)$ can depend on z . Letting $J \rightarrow \infty$, we conclude that $f(z) - g(z)$ converges in probability to zero for any fixed $z \in \Omega$.

Given any smooth compactly supported function $F : \Omega \rightarrow \mathbb{C}$, define the random variables

$$F(\Lambda_f) := \sum_{w \in \Lambda_f} F(w)$$

and

$$F(\Lambda_g) := \sum_{w \in \Lambda_g} F(w),$$

where Λ_f, Λ_g are the zeroes of f, g respectively (counting multiplicity). By the Stone-Weierstrass theorem (using (19), (20) to control errors that are small in the uniform norm), it suffices to show that

$$\mathbf{E} F_1(\Lambda_f) \dots F_k(\Lambda_f) = \mathbf{E} F_1(\Lambda_g) \dots F_k(\Lambda_g) + o(1)$$

for all smooth compactly supported $F_1, \dots, F_k : \Omega \rightarrow \mathbb{C}$. From part (i) of Theorem 1.11, we know that $F_1(\Lambda_f) \dots F_k(\Lambda_f)$ is uniformly integrable in n (indeed, it has bounded L^m norm for each m). Thus it suffices to show that $F_1(\Lambda_f) \dots F_k(\Lambda_f) - F_1(\Lambda_g) \dots F_k(\Lambda_g)$ converges in probability to zero. By another appeal to Theorem 1.11(i), it suffices to show that $F_j(\Lambda_f) - F_j(\Lambda_g)$ converges in probability to zero for each j .

Fix j , and write F for F_j . By Green's theorem, we can write

$$F(\Lambda_f) = \frac{1}{2\pi} \int_{\mathbb{C}} (\Delta F(z)) \log |f(z)| d^2 z$$

where $\Delta := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the usual Laplacian. Similarly for $F(\Lambda_g)$. Thus it suffices to show that

$$\int_{\mathbb{C}} (\Delta F(z)) (\log |f(z)| - \log |g(z)|) d^2 z$$

converges in probability to zero.

We already know that for fixed z , $f(z) - g(z)$ converges in probability to zero. The function ΔF is bounded and compactly supported. Also, by Lemma To conclude the claim it then suffices by a truncation argument (cf. [44, Lemma 3.1]) to obtain the uniform integrability bounds

$$\mathbf{E} \int_{\text{supp}(F)} |\log |f(z)||^2 + |\log |g(z)||^2 d^2 z = O(1).$$

But the bound for f follows from (15); the bound for g can be deduced from f by a Fatou lemma type argument. The proof of Theorem 1.11 is now complete.

6. ZERO ROW SUM

We now prove Theorem 1.13. We begin by proving the spectral radius upper bound

$$\rho\left(\frac{1}{\sqrt{n}} X_n P_n\right) \leq 1 + o(1)$$

which holds almost surely. It will suffice to show that almost surely one has

$$\left\| \left(\frac{1}{\sqrt{n}} X_n P_n\right)^m \right\|_{op} \leq O(m^{O(1)}) + o(1)$$

for each $m \geq 1$. Writing $P_n = 1 - (1 - P_n)$, expanding, and applying Theorem 1.4 and the fact that the operator norm forms a Banach algebra, it then suffices to show that

$$\|(1 - P_n) \left(\frac{1}{\sqrt{n}} X_n\right)^j (1 - P_n)\|_{op} = o(1)$$

almost surely for each fixed j , as this handles all but the $O(m)$ terms in the expansion that involve at most one factor of $1 - P_n$, each of which is $O(m^2)$ at worst by Theorem 1.4. But this bound follows from Lemma 2.3.

The spectral radius lower bound will follow from the circular law claim. Since almost sure convergence implies convergence in probability by the dominated convergence theorem, it will suffice to show that $\mu_{\frac{1}{\sqrt{n}} X_n P_n}$ and $\mu_{\frac{1}{\sqrt{n}} X_n}$ have the same almost sure limit. Applying the replacement principle ([44, Theorem 2.1]), it suffices to show that for almost every complex number z , one has

$$(22) \quad \frac{1}{n} \log |\det(\frac{1}{\sqrt{n}} X_n P_n - z)| - \frac{1}{n} \log |\det(\frac{1}{\sqrt{n}} X_n - z)|$$

converges almost surely to zero.

Fix z ; we may take z to be non-zero. We allow implied constants in the $O()$ notation to depend on z . We can rewrite (22) as

$$(23) \quad \frac{1}{2} \left(\int_0^\infty \log t d\nu'_n(t) - \int_0^\infty \log t d\nu_n(t) \right)$$

where ν'_n, ν_n are the ESDs of $(\frac{1}{\sqrt{n}} X_n P_n - z)(\frac{1}{\sqrt{n}} X_n P_n - z)^*$ and $(\frac{1}{\sqrt{n}} X_n - z)(\frac{1}{\sqrt{n}} X_n - z)^*$ respectively.

The matrix $\frac{1}{\sqrt{n}}X_nP_n - z$ is a rank one perturbation of $\frac{1}{\sqrt{n}}X_n - z$, and so the singular values $\sigma_i(\frac{1}{\sqrt{n}}X_nP_n - z)$ of the former interlace that of the latter in the sense that

$$(24) \quad \sigma_{i-1}\left(\frac{1}{\sqrt{n}}X_n - z\right) \geq \sigma_i\left(\frac{1}{\sqrt{n}}X_nP_n - z\right) \geq \sigma_{i+1}\left(\frac{1}{\sqrt{n}}X_nP_n - z\right)$$

whenever i is such that the expressions are well-defined (i.e. $1 < i \leq n$ for the first inequality and $1 \leq i < n$ for the second). This adequately controls all of the singular values of $\frac{1}{\sqrt{n}}X_nP_n - z$ except for the smallest and largest. But from Theorem 1.4 we know that the largest singular value of both matrices are $O(1)$. From [44, Lemma 4.1] we almost surely also have a lower bound

$$\sigma_n\left(\frac{1}{\sqrt{n}}X_n - z\right) \gg n^{-O(1)}$$

for all sufficiently large n . So if we can also obtain the corresponding bound

$$(25) \quad \sigma_n\left(\frac{1}{\sqrt{n}}X_nP_n - z\right) \gg n^{-O(1)}$$

almost surely for all sufficiently large n , then by the alternating series test⁶ we see that

$$(23) = O\left(\frac{1 + |\log n^{-O(1)}|}{n}\right) = o(1)$$

as required. So it will suffice to establish the least singular value bound (25). By the Borel-Cantelli lemma, it will suffice to show that

$$\mathbf{P}(\sigma_n\left(\frac{1}{\sqrt{n}}X_nP_n - z\right) \leq n^{-C}) = O(n^{-2})$$

(say) for all sufficiently large n , and some absolute constant C . Taking transposes, it suffices to show that

$$\mathbf{P}(\sigma_n\left(\frac{1}{\sqrt{n}}P_nX_n - z\right) \leq n^{-C}) = O(n^{-2}).$$

Let C be chosen later. In order for the above event to hold, there must exist a unit vector v such that

$$\left\|\frac{1}{\sqrt{n}}P_nX_nv - zv\right\| \leq n^{-C}.$$

We now work to eliminate the role of the projection P_n by dropping a dimension. Taking inner products with ϕ_n , we see that

$$z|\langle v, \phi_n \rangle| \leq n^{-C}.$$

Since z is fixed and non-zero, we thus see that

$$v - P_nv = O(n^{-C+O(1)})$$

and thus

$$\frac{1}{\sqrt{n}}P_nX_nP_nv - zP_nv = O(n^{-C+O(1)}).$$

If we let $v' := v - P_nv/\|v - P_nv\|$, we thus see that v' is orthogonal to ϕ_n and

$$\frac{1}{\sqrt{n}}P_nX_nv' - zv' = O(n^{-C+O(1)})$$

and thus

$$\frac{1}{\sqrt{n}}X_nv' - zv' = \alpha\phi_n + O(n^{-C+O(1)})$$

for some complex number α . Writing

$$v' := (v'_1, \dots, v'_{n-1}, -v'_1 - \dots - v'_n)$$

⁶More precisely, the integrals $\int_0^\infty \log t \, d\nu_n(t)$ and $\int_0^\infty \log t \, d\nu'_n(t)$ are both averages of n increasing quantities of size $O(1 + |\log n^{-O(1)}|)$; by the interlacing property (24) (which bounds the even terms in the latter average by the odd terms in the former, and vice versa), the difference between these two averages can be rearranged as an average of two alternating series whose terms are increasing in magnitude.

we thus have

$$\sum_{j=1}^{n-1} \left[\frac{1}{\sqrt{n}} (x_{ij} - x_{in}) - z(\delta_{ij} - \delta_{in}) \right] v'_j = \frac{\alpha}{\sqrt{n}} + O(n^{-C+O(1)})$$

for all $i = 1, \dots, n$. Subtracting off the $i = n$ equation to eliminate α , we conclude that

$$\sum_{j=1}^{n-1} \left[\frac{1}{\sqrt{n}} (x_{ij} - x_{in} - x_{nj} + x_{nn}) - z(\delta_{ij} + 1) \right] v'_j = O(n^{-C+O(1)})$$

for all $i = 1, \dots, n-1$. Since v' has unit norm, (v'_1, \dots, v'_{n-1}) has norm between 1 and $1/2\sqrt{n}$, and we conclude that

$$(26) \quad \sigma_{n-1} \left(\frac{1}{\sqrt{n}} X_{n-1} + D_{n-1} \right) = O(n^{-C+O(1)})$$

where X_{n-1}, D_{n-1} are the $n-1 \times n-1$ matrices

$$X_{n-1} := (x_{ij})_{1 \leq i, j \leq n-1}$$

and

$$D_{n-1} := \left(\frac{1}{\sqrt{n}} (-x_{in} - x_{nj} + x_{nn}) - z\delta_{ij} + z \right)_{1 \leq i, j \leq n-1}.$$

If we condition x_{in}, x_{jn}, x_{nn} to be fixed, then D_{n-1} is deterministic, while X_{n-1} remains an iid random matrix. Applying [44, Lemma 4.1], [42, Theorem 2.1], or [43, Theorem 4.1], we see that the conditional probability of (26) is $O(n^{-2})$ if C is large enough, and if the x_{in}, x_{nj}, x_{nn} are bounded by (say) n^{100} in magnitude. Integrating out the conditioning (and using Chebyshev's inequality and the union bound to handle the rare event when one of the entries x_{in}, x_{nj}, x_{nn} is larger than n^{100} in magnitude) we obtain the claim. This concludes the proof of Theorem 1.13.

REFERENCES

- [1] N. Alon, M. Krivelevich, V.H. Vu, *On the concentration of eigenvalues of random symmetric matrices*, Israel J. Math. **131** (2002) 259-267.
- [2] A. Andrew, *Eigenvalues and singular values of certain random matrices*, J. Comput. Appl. Math. **30** (1990), no. 2, 165-171.
- [3] P. Arbenz, W. Gander, G. Golub, *Restricted rank modification of the symmetric eigenvalue problem: Theoretical considerations*, Linear Algebra and its Applications **104** (1988), 75-95.
- [4] Z. D. Bai, *Circular law*, Ann. Probab. **25** (1997), 494-529.
- [5] Z. D. Bai, J. Silverstein, *Exact separation of eigenvalues of large-dimensional sample covariance matrices*, Ann. Probab. **27** (1999), no. 3, 1536-1555.
- [6] Z. D. Bai and J. Silverstein, *Spectral analysis of large dimensional random matrices*, Mathematics Monograph Series **2**, Science Press, Beijing 2006.
- [7] J. Baik, J. Silverstein, *Eigenvalues of large sample covariance matrices of spiked population models*. (English summary) J. Multivariate Anal. **97** (2006), no. 6, 1382-1408.
- [8] D. Bai, J.-F. Yao, *Central limit theorems for eigenvalues in a spiked population model*, Ann. I.H.P.-Prob. et Stat., 2008, Vol. 44 No. 3, 447-474.
- [9] Z. D. Bai, Y. Q. Yin, *Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix*, Ann. Probab. **16** (1988), no. 4, 1729-1741.
- [10] Z. D. Bai, Y. Q. Yin, *Limiting behavior of the norm of products of random matrices and two problems of Geman-Hwang*, Probab. Theory Relat. Fields **73** (1986), 555-569.
- [11] J. Baik, G. Ben Arous, S. Peche, *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*, Ann. Probab. **33** (2005), no. 5, 1643-1697.
- [12] J. Baik, J. Silverstein, *Eigenvalues of large sample covariance matrices of spiked population models*, J. Multivariate Anal. **97** (2006), no. 6, 1382-1408.
- [13] F. Benaych-Georges, R. Rao, *The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices*, preprint.
- [14] F. Benaych-Georges, A. Guionnet, M. Maida, *Large deviations of the extreme eigenvalues of random deformations of matrices*, preprint.
- [15] F. Benaych-Georges, A. Guionnet, M. Maida, *Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices*, preprint.
- [16] C. Bordenave, *On the spectrum of sum and product of non-hermitian random matrices*, preprint. Available at <http://arxiv.org/abs/1010.3087>.

- [17] C. Bordenave, P. Caputo, D. Chafaï, *Circular law theorem for random Markov matrices*, preprint. Available at <http://arxiv.org/abs/0808.1502>.
- [18] M. Capitaine, C. Donati-Martin, D. Féral, *Central limit theorems for eigenvalues of deformations of Wigner matrices*, preprint. Available at <http://arxiv.org/abs/0903.4740>.
- [19] D. Chafaï, *Circular law for non-central random matrices*, J. Th. Prob. **23** (2010) 945–950.
- [20] W. Feller, *An introduction to probability theory and its applications*, volume II. John Wiley & Sons, Inc., New York-London-Sydney, 1971.
- [21] D. Féral, S. Péché, *The largest eigenvalue of rank one deformation of large Wigner matrices*, Commun. Math. Phys. **272** (2007), 185–228.
- [22] S. Geman, *The spectral radius of large random matrices*, Ann. Probab. **14** (1986), 1318–1328.
- [23] J. Ginibre, *Statistical Ensembles of Complex, Quaternion, and Real Matrices*, Journal of Mathematical Physics **6** (1965), 440–449.
- [24] V. L. Girko, *Circular law*, Theory Probab. Appl. (1984), 694–706.
- [25] V. L. Girko, *The strong circular law. Twenty years later. II*, Random Oper. Stochastic Equations **12** (2004), no. 3, 255–312.
- [26] F. Götze, A.N. Tikhomirov, *On the circular law*. Available at arxiv.org/abs/math/0702386.
- [27] F. Götze, A.N. Tikhomirov, *The Circular Law for Random Matrices*, Ann. Probab. **38** (2010), no. 4, 1444–1491.
- [28] A. Guionnet, O. Zeitouni, *Concentration of the spectral measure for large matrices*, Electron. Comm. Probab. **5** (2000) 119–136.
- [29] B. J. Hough, M. Krishnapur, Y. Peres, B. Virág, *Zeros of Gaussian analytic functions and determinantal point processes*. University Lecture Series, 51. American Mathematical Society, Providence, RI, 2009.
- [30] M. Ledoux, *The Concentration of Measure Phenomenon*. American Mathematical Society, 2001.
- [31] A. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*, Adv. Math. **195** (2005), no. 2, 491–523.
- [32] M. Meckes, *Concentration of norms and eigenvalues of random matrices*, J. Funct. Anal. **211** (2004), no. 2, 508–524.
- [33] M.L. Mehta, *Random Matrices and the Statistical Theory of Energy Levels*, Academic Press, New York, NY, 1967.
- [34] G. Pan and W. Zhou, *Circular law, Extreme singular values and potential theory*, J. Multivariate Anal. **101** (2010), no. 3, 645–656.
- [35] S. Peche, *The largest eigenvalue of small rank perturbations of Hermitian random matrices*, Probab. Theory Related Fields **134** (2006), no. 1, 127–173.
- [36] Y. Peres, B. Virág, *Zeros of the i.i.d. Gaussian power series: a conformally invariant determinantal process*, Acta Math. **194** (2005), no. 1, 135.
- [37] A. Pizzo, D. Renfrew, A. Soshnikov, *On finite rank deformations of Wigner matrices*, preprint.
- [38] T. Prosen, *Exact statistics of complex zeros for Gaussian random polynomials with real coefficients*, J. Phys. A: Math. Gen. **29** (1996) 4417–4423.
- [39] K. Rajan, L. F. Abbott, *Eigenvalue Spectra of Random Matrices for Neural Networks*, Phys. Rev. Lett. **97** (2006), 188104.
- [40] M. Rudelson, R. Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), no. 2, 600–633.
- [41] J. Silverstein, *The spectral radii and norms of large-dimensional non-central random matrices*, Comm. Statist. Stochastic Models **10** (1994), no. 3, 525–532.
- [42] T. Tao and V. Vu, *Random Matrices: The circular Law*, Communications in Contemporary Mathematics, **10** (2008), 261–307.
- [43] T. Tao and V. Vu, *Smooth analysis of the condition number and the least singular value*, preprint. Available at <http://arxiv.org/abs/0805.3167>.
- [44] T. Tao, V. Vu, M. Krishnapur, *Random matrices: Universality of ESDs and the circular law*, Ann. Probab. **38** (2010), no. 5, 2023–2065.

DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES CA 90095-1555

E-mail address: tao@math.ucla.edu